carbon

Artificial Intelligence for Financial Inclusion.

Alternative Data and Risk Prediction.

Graph Databases: A key to addressing Financial Services challenges.

I - Introduction, Definitions

- Credit Scoring
  - Traditional methods
    - Credit Bureau
  - Alternative Data
    - Unbanked population
    - Smartphone and Social Networks : A gold mine?

II - Methods

- Payment Default (PD) & Risk Prediction
- Algorithms selection overview:
  - Naive
  - SVM
  - Random Forest
  - Extreme and Light Gradient Boosting
  - Deep Learning

App I : Why and When XGB is better than DP

III - Graph Database - Neo4J

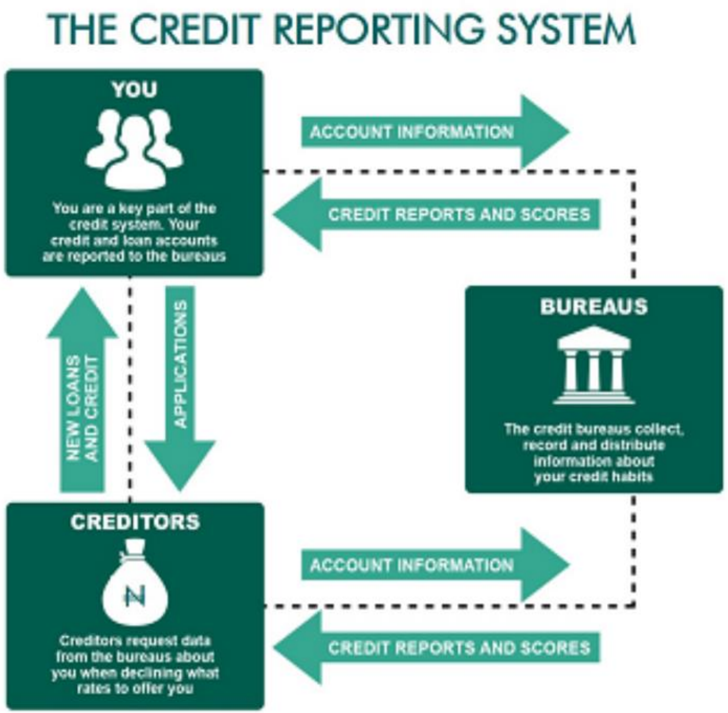
- Introduction
  - Difference with other databases?
  - Why is it useful?
- Neo4j
  - Structure
  - Cypher language
  - Graph Algorithms
- Graph DB at Carbon

IV - Carbon use cases

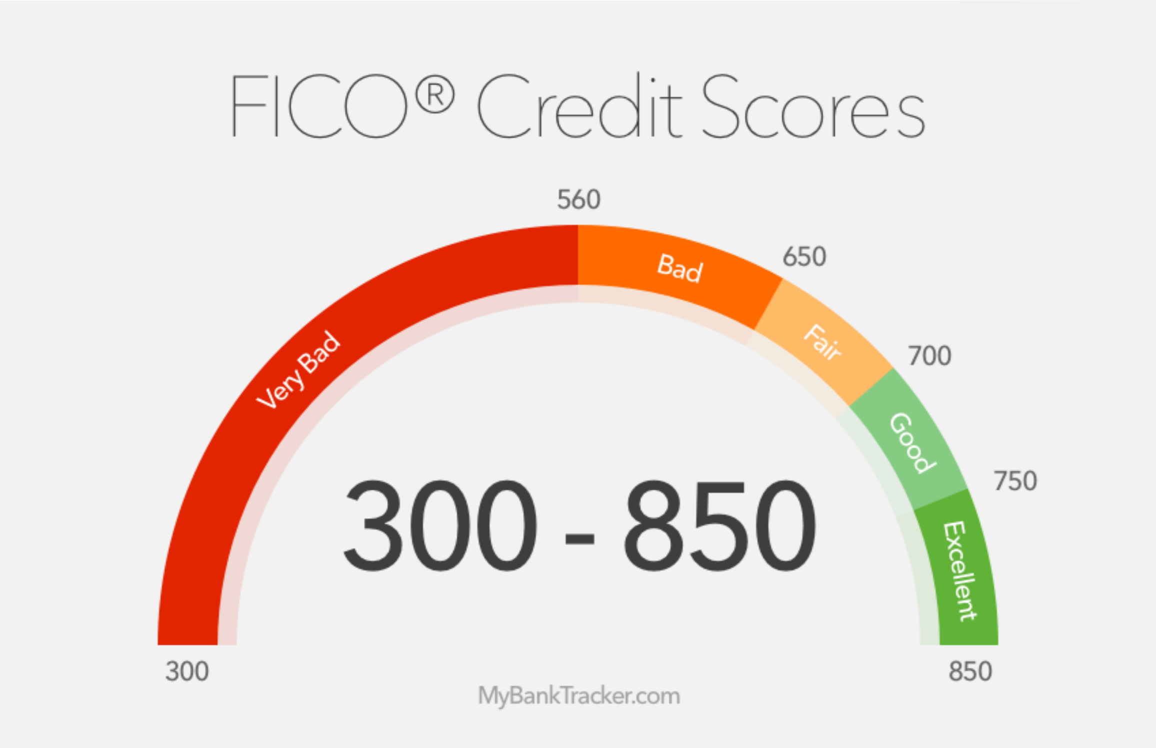
@ Carbon

- How is CARBON using ML for credit scoring and risk prediction
  - DataRobot
  - CARBON ML Architecture & Pipelines overview
- AI Bias vs Human Logical Error
- Conclusion

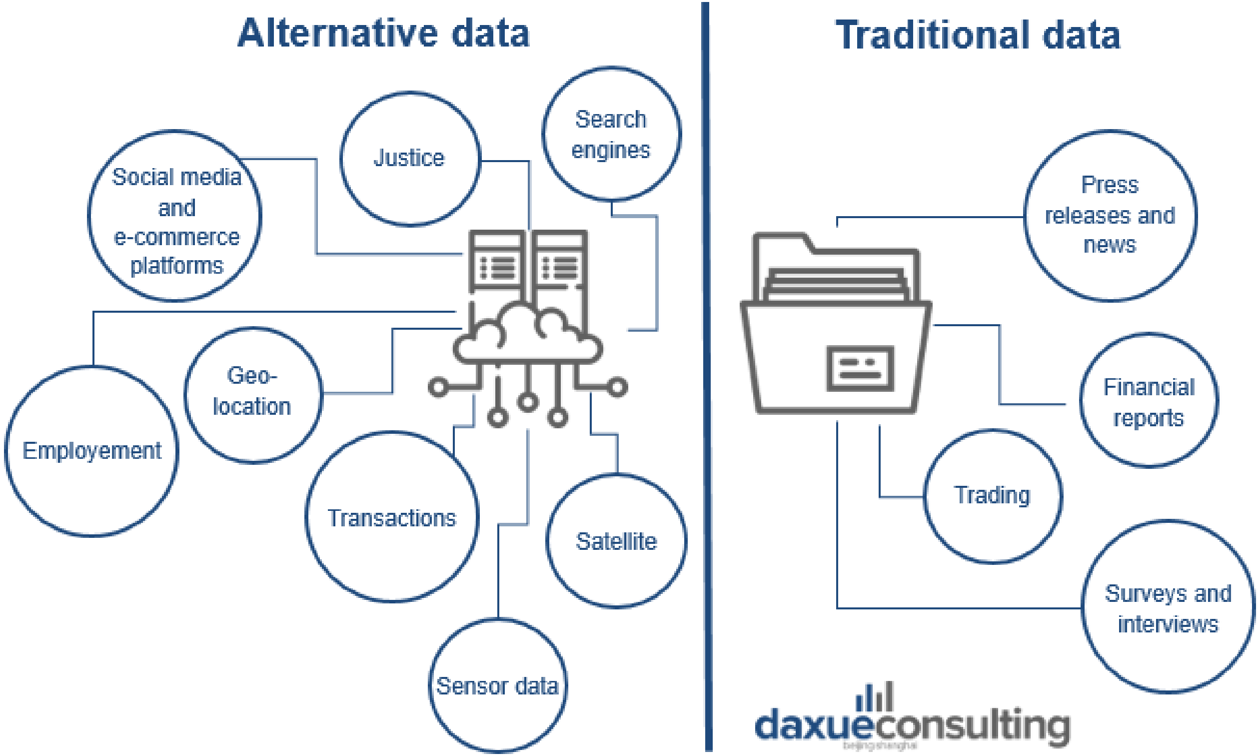
Credit Scoring - Traditional methods



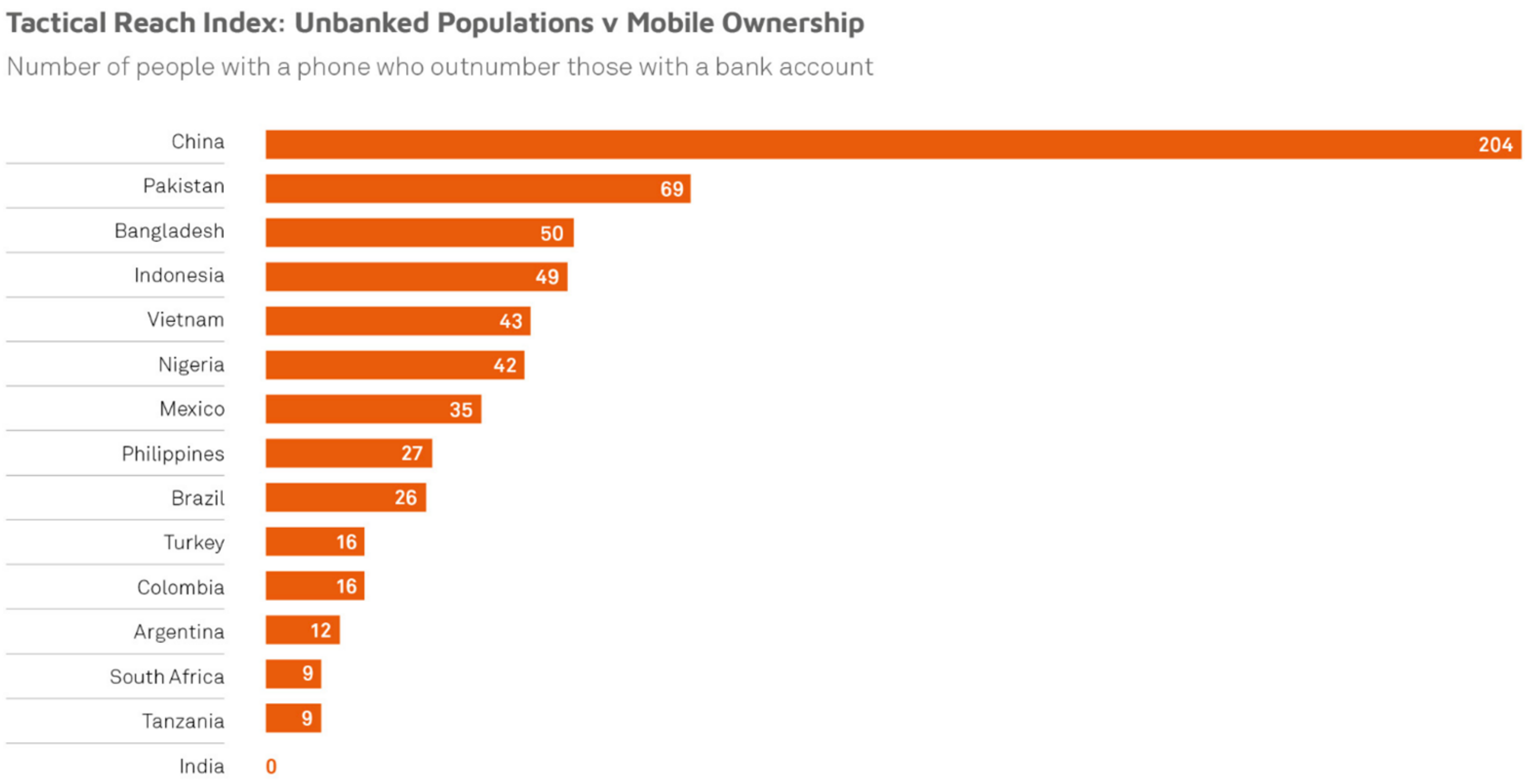
Credit Scoring - Traditional methods



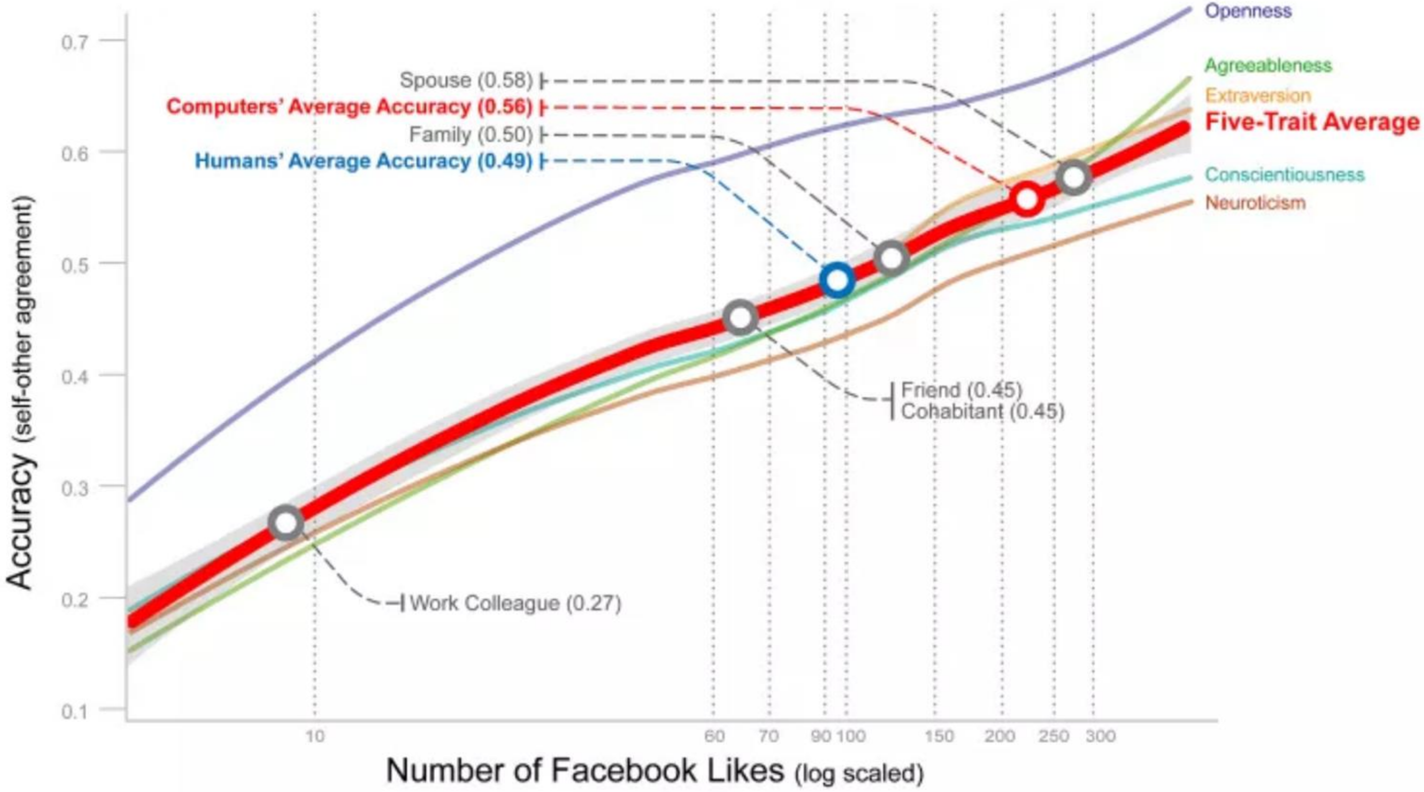
Alternative Data



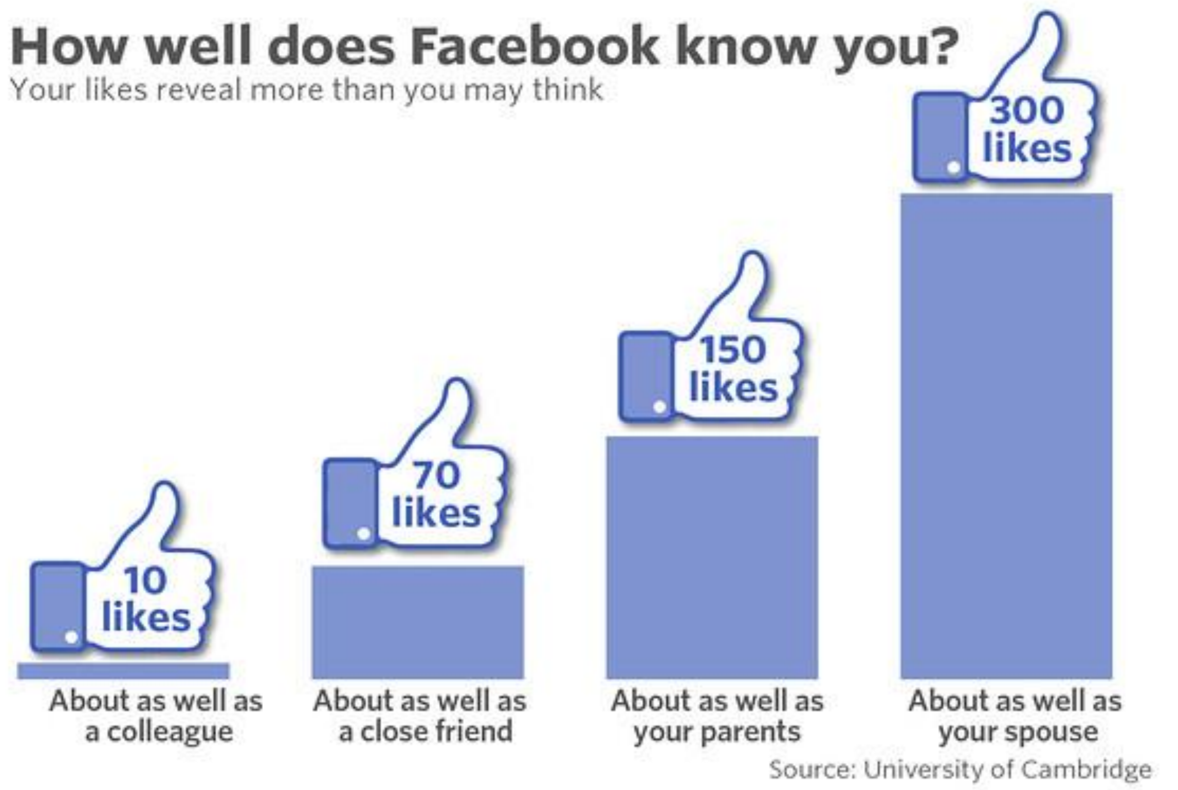
Unbanked population



Smartphone and Social Networks : A gold mine?



Smartphone and Social Networks : A gold mine?



carbon

Methods

- Payment Default (PD) & Risk Prediction
- Algorithms selection overview

Payment Default (PD) & Risk Prediction

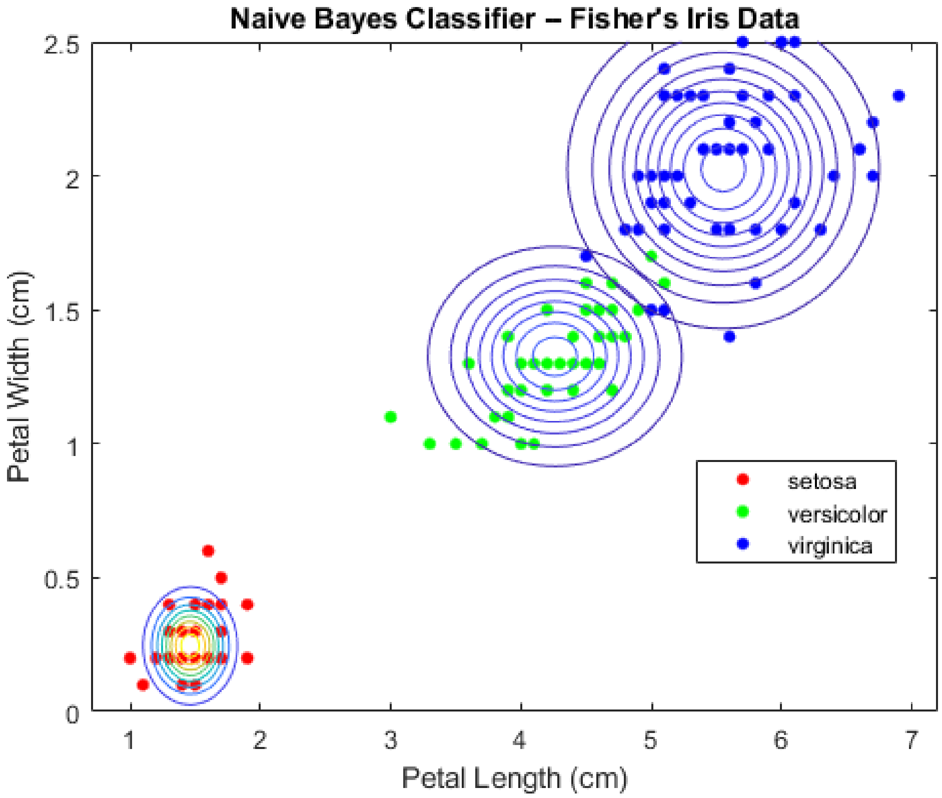
Likelihood \ Impact	Negligible	Minor	Moderate	Significant	Severe
Very Likely	Low Med	Medium	Med Hi	High	High
Likely	Low	Low Med	Medium	Med Hi	High
Possible	Low	Low Med	Medium	Med Hi	Med Hi
Unlikely	Low	Low Med	Low Med	Medium	Med Hi
Very Unlikely	Low	Low	Low Med	Medium	Medium

Risk Assessment Matrix

Algorithms selection overview

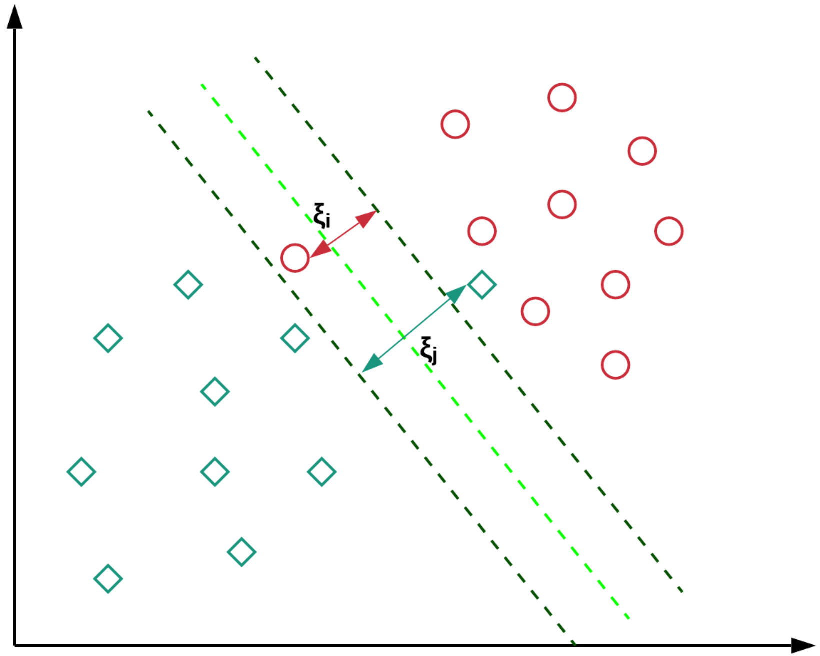
Naive Bayes

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$



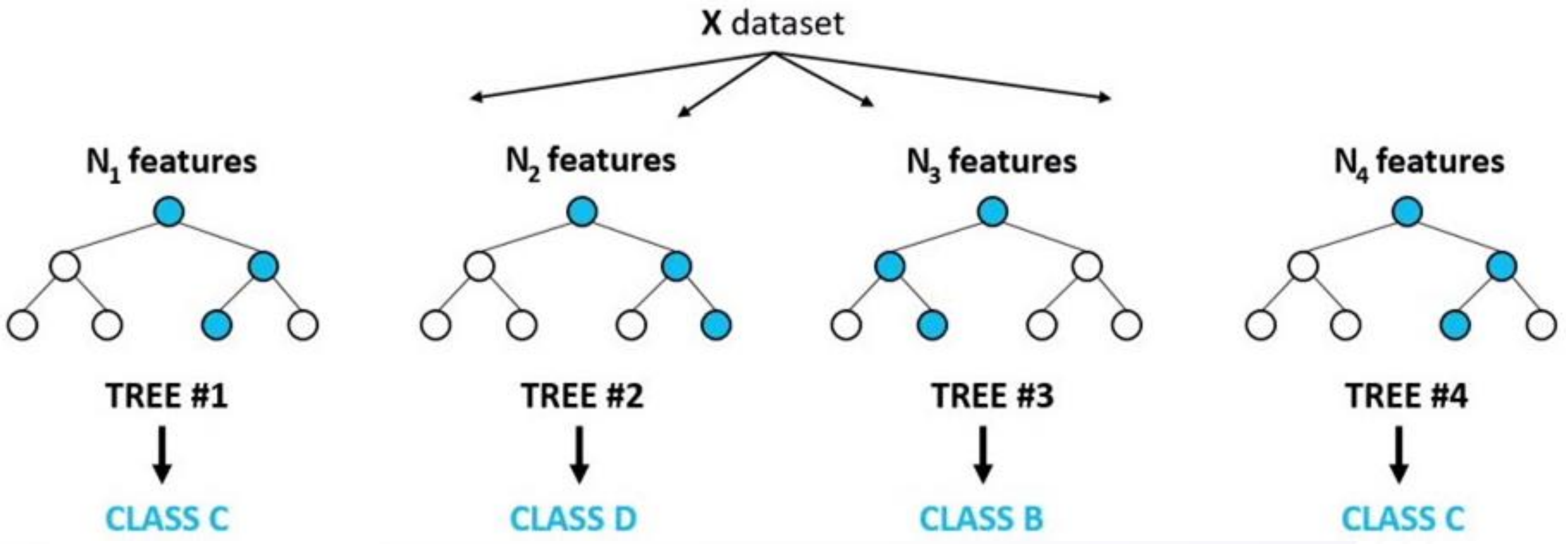
Algorithms selection overview

SVM
(Support Vector Machines)



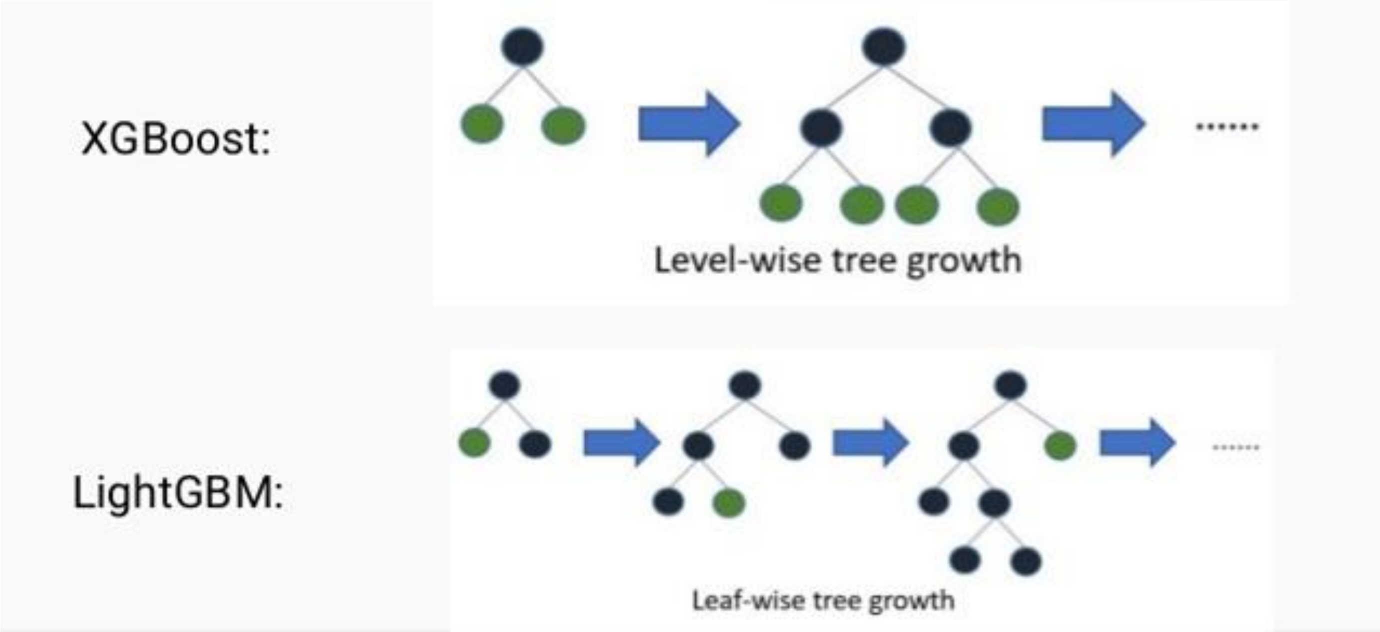
Algorithms selection overview

Random Forest



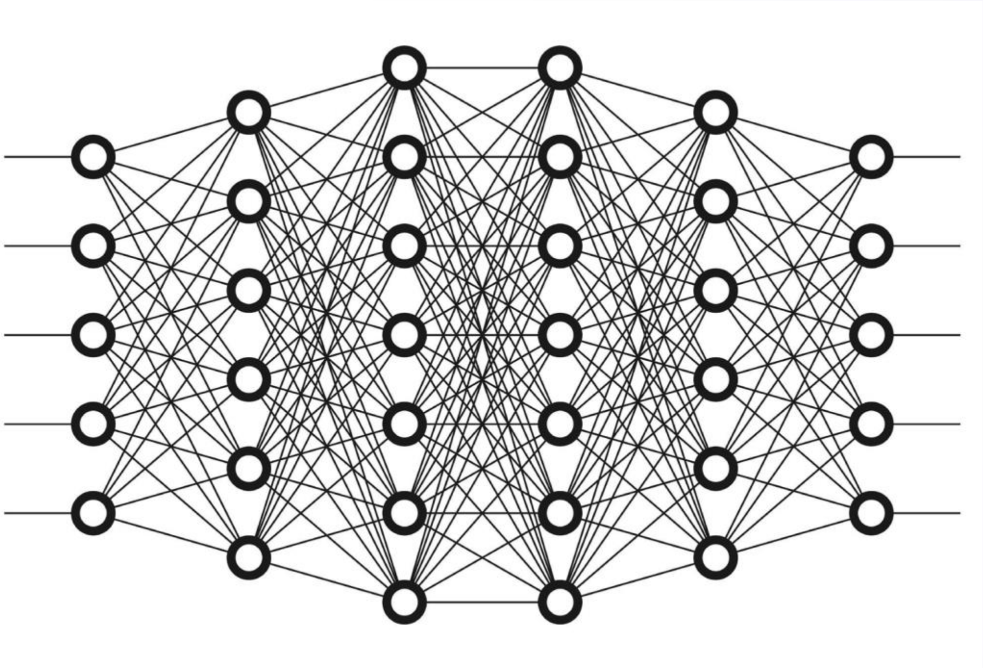
Algorithms selection overview

Extreme and Light Gradient Boosting



Algorithms selection overview

Deep Learning



Why and When XGB is better than DP?

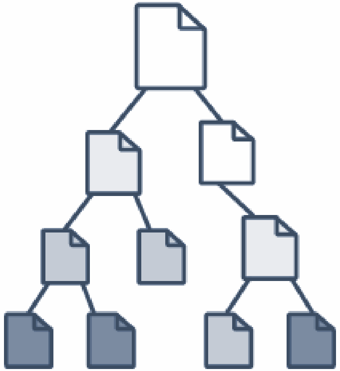
carbon

Graph Databases

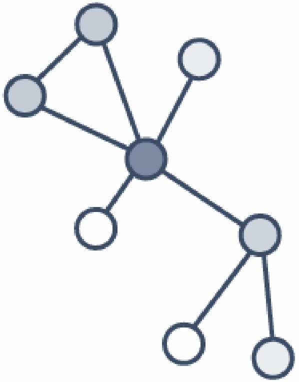
A key to addressing Financial Services challenges.

Types of databases

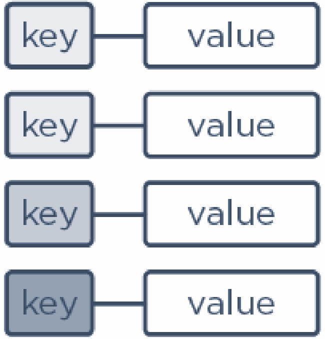
Document



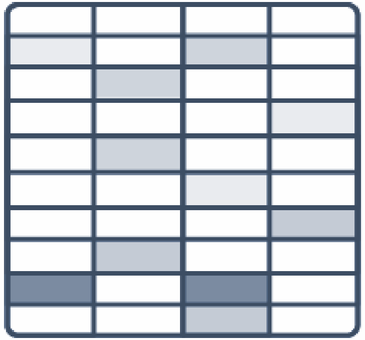
Graph



Key-Value



Wide-column



NO GOOD, NO BAD

Example: Library

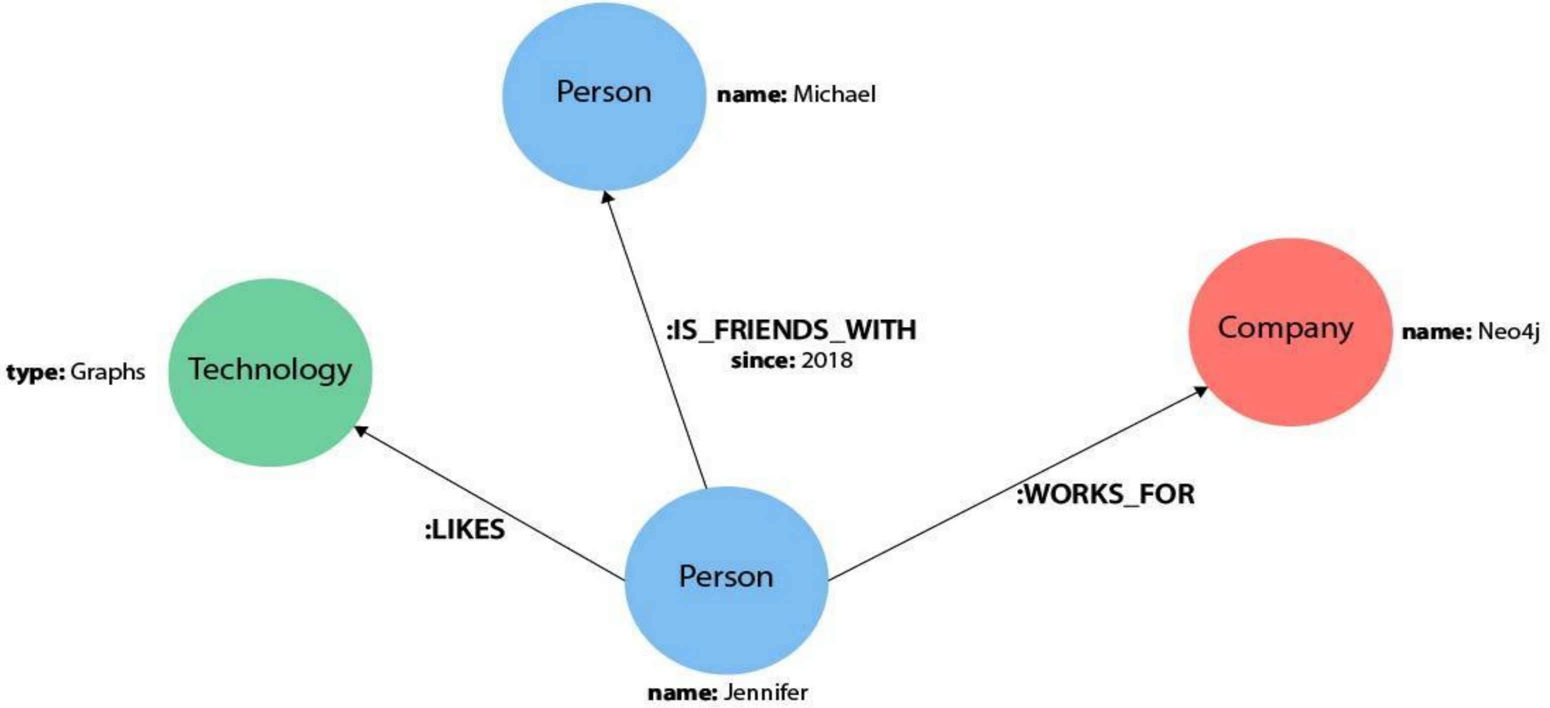
NoSQL / Document	SQL / Wide-column	Key-Value	Graph
Book content	Author/Publisher/Date	Book availability	Book usage

Why Graph DB?

Focus on communities and not individuals.
Behaviour is highly influenced by the community.

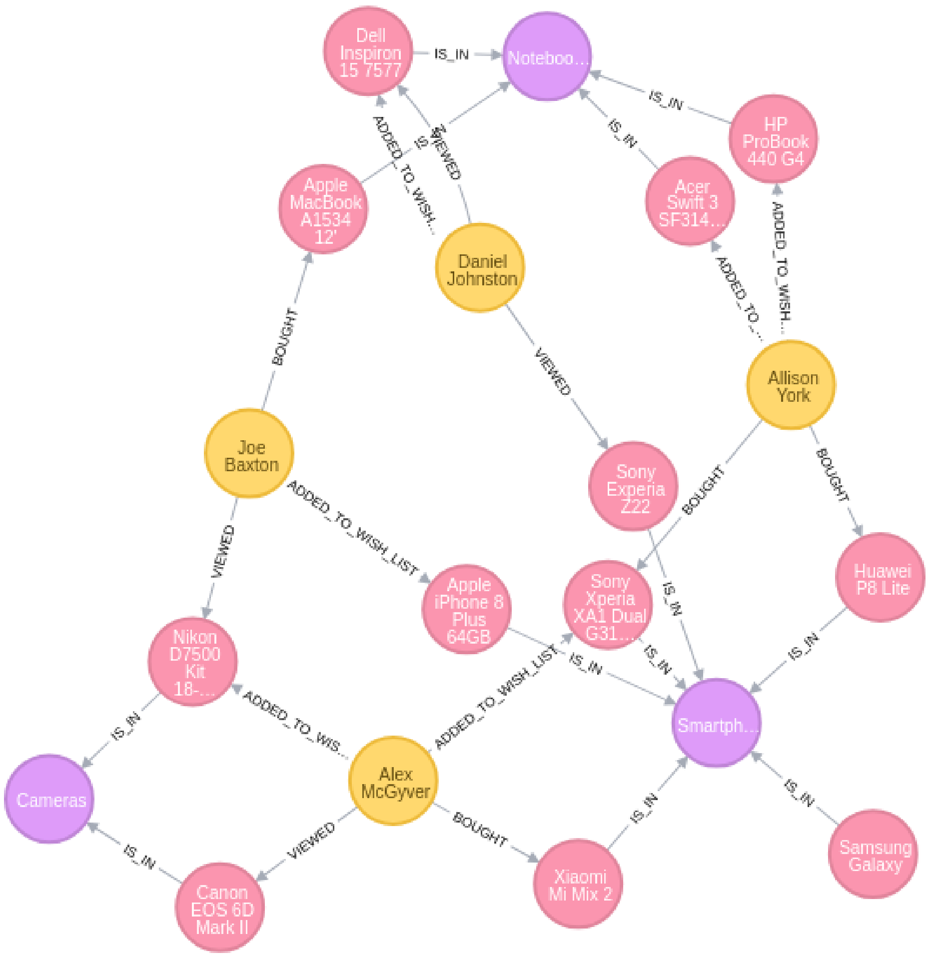
Understanding the customers:
Do they know each other? → Fraud detection
Should they know each other? → Recommendation / Marketing

Neo4j



Person
name: Michael
:IS_FRIENDS_WITH
since: 2018
Company
name: Neo4j
type: Graphs
Technology
:LIKES
:WORKS_FOR
Person
name: Jennifer

Neo4j


Dell Inspiron 15 7577
Noteboo...
HP ProBook 440 G4
Acer Swift 3 SF314...
Apple MacBook A1534 12'
Daniel Johnston
Allison York
Joe Baxton
Sony Experia Z22
Sony Xperia XA1 Dual G31...
Huawei P8 Lite
Apple iPhone 8 Plus 64GB
Nikon D7500 Kit 18-...
Smartph...
Alex McGyver
Cameras
Canon EOS 6D Mark II
Xiaomi Mi Mix 2
Samsung Galaxy
IS_IN
VIEWED
BOUGHT
ADDED_TO_WISH_LIST
ADDED_TO_WISH...
ADDED_TO_...
ADDED_TO_WIS...

Cypher

carbon

```
//data stored with this direction
 CREATE (e:Employee)-[:WORKS_AT]->(c:Company)
```

Cypher

```
//data stored with this direction
 CREATE (e:Employee {name:'Jacobo'})-[:WORKS_AT]->(c:Company {name:'Carbon'})
```

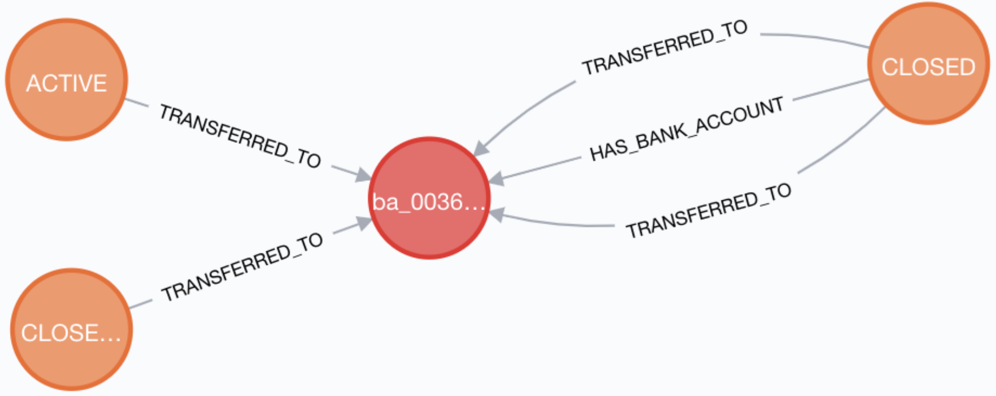
Cypher

carbon

```
//retrieve names of Carbon employees
MATCH (e:Employee)-[:WORKS_AT]->(c:Company {name:'Carbon'})
RETURN e.name
```

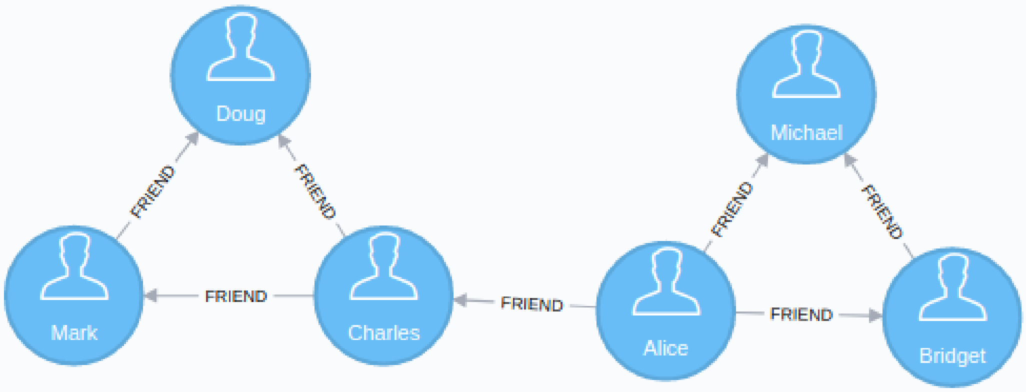
Example: bank transfers

```
MATCH p = (c:client)-->(:bank_account)<--(c2:client)
 WHERE c<>c2
 RETURN p
```

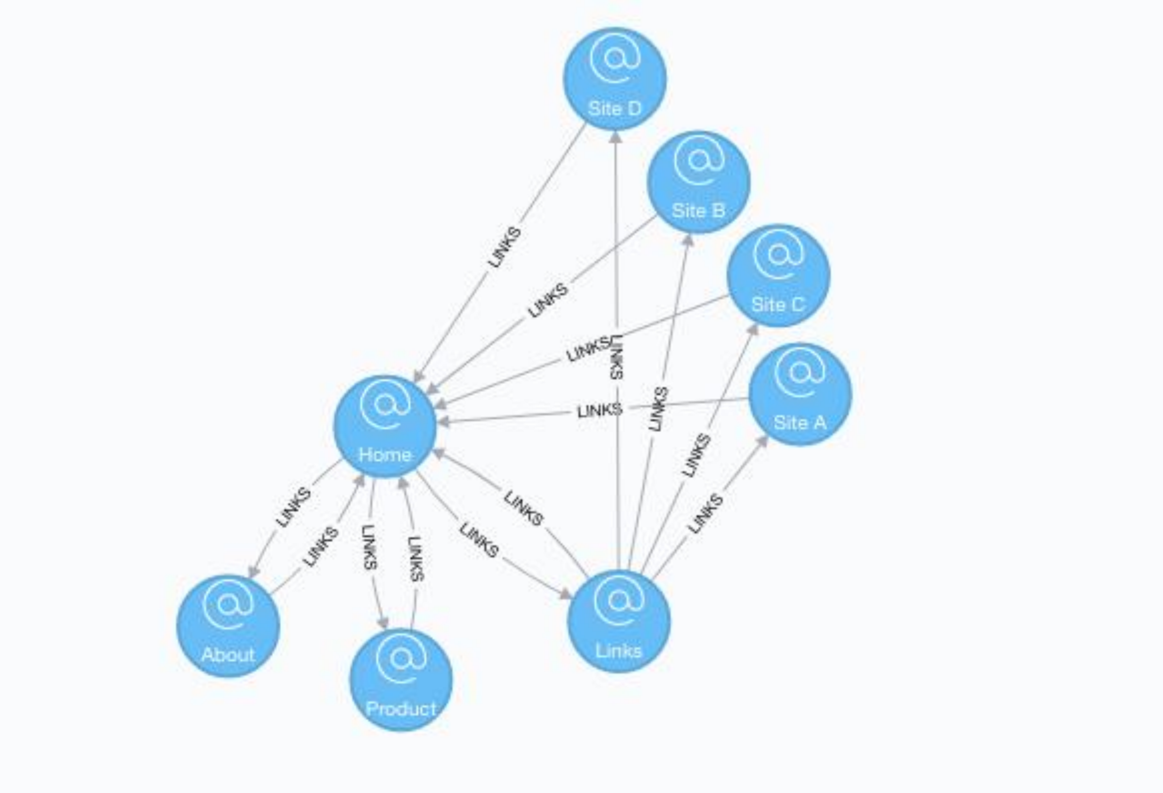


Graph Algorithms

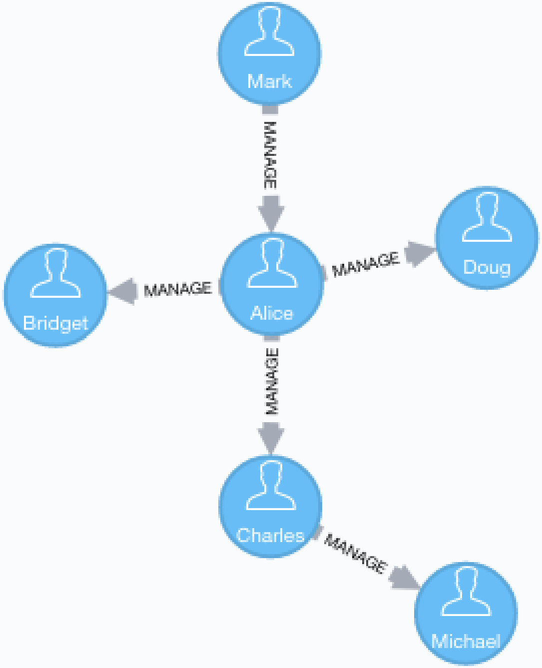
Community detection



Centrality: PageRank



Centrality: Betweenness



Graph DB at Carbon

carbon

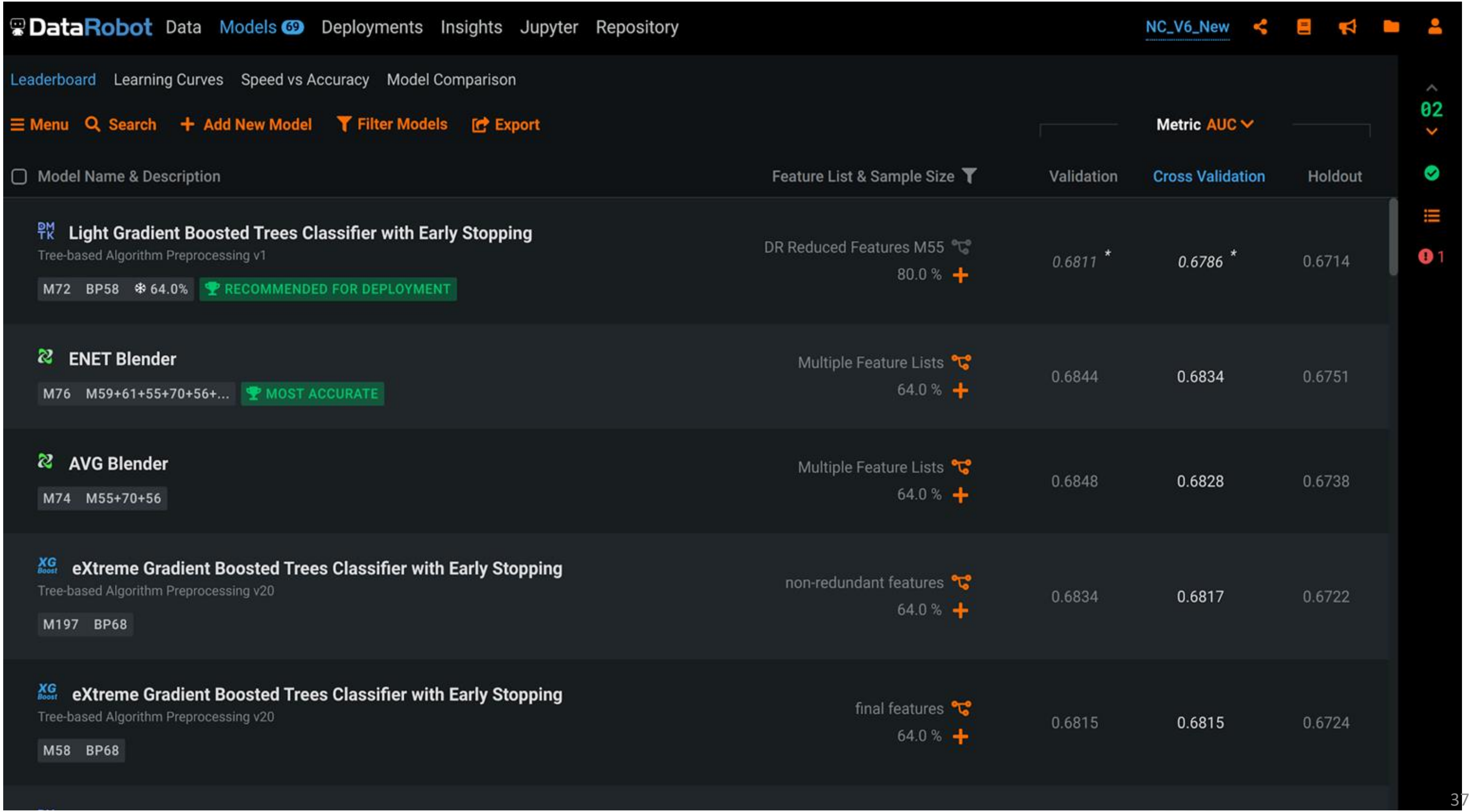
- Marketing campaigns
- Fraud prevention

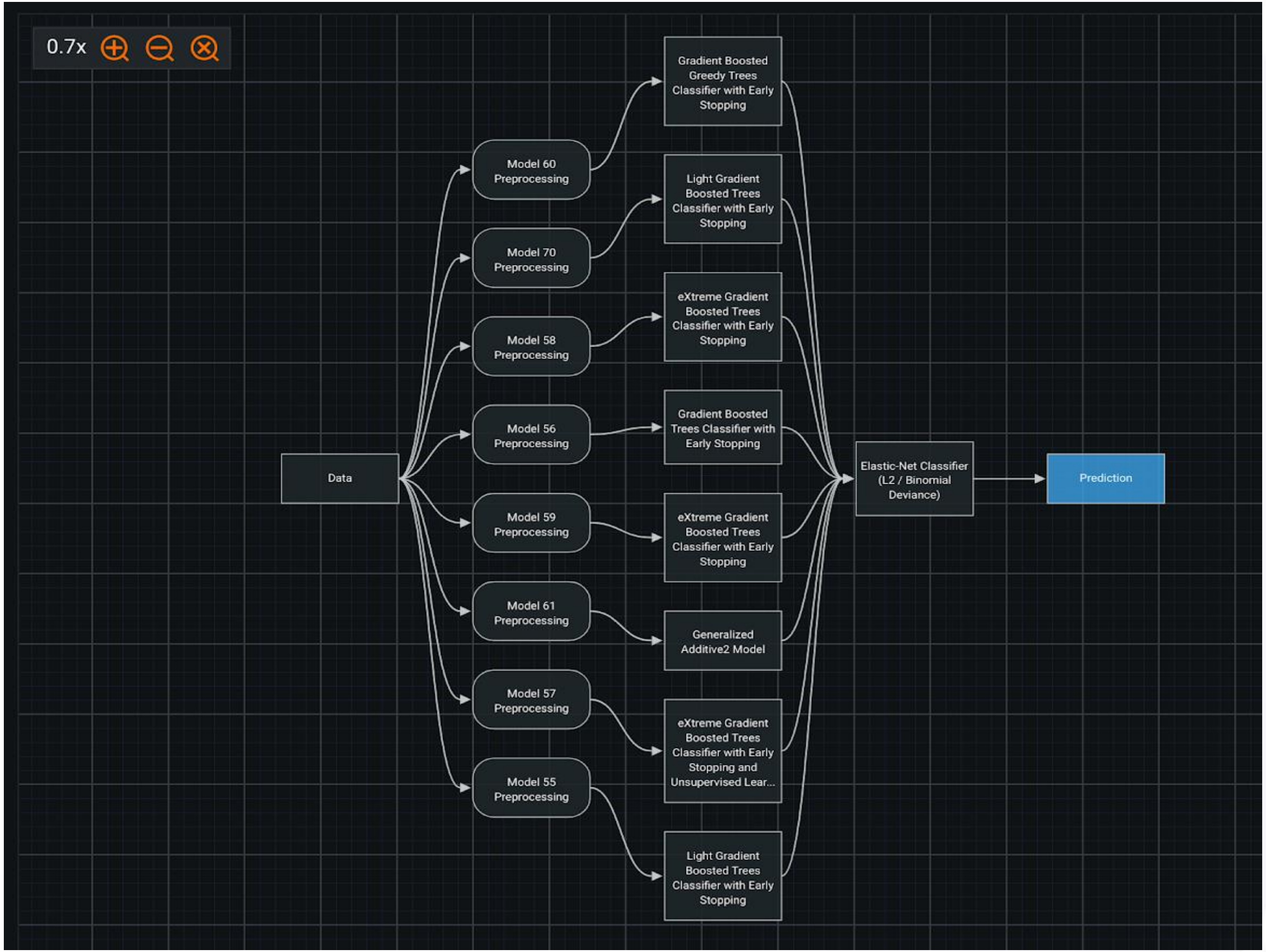
Carbon use cases

1. How are we using Machine Learning at CARBON for credit scoring
   and risk prediction?

DataRobot

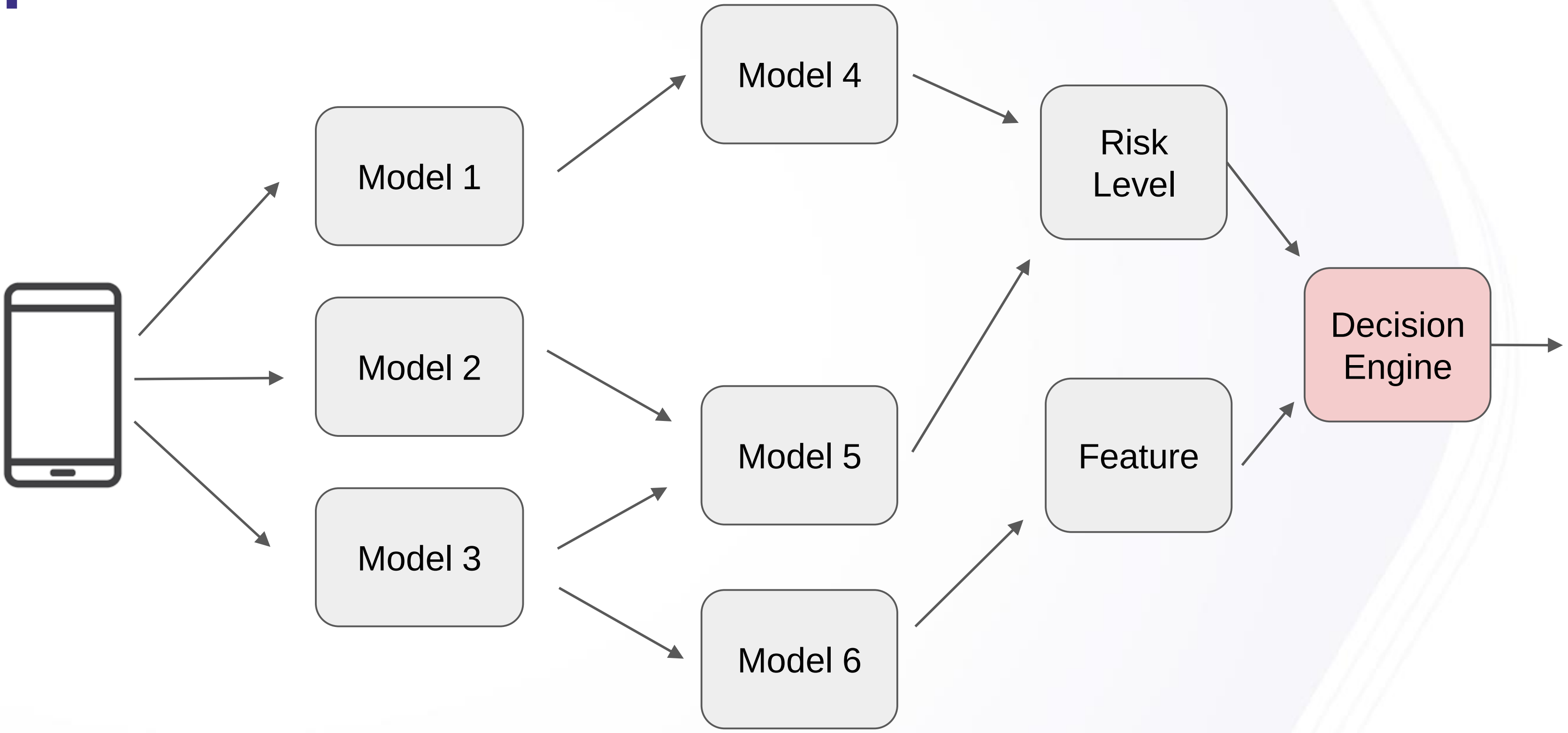
DataRobot Data Models Deployments Insights Jupyter Repository
What would you like to predict?
Sales REGRESSION
Number of rows
Sales
Start
Modeling Mode: Autopilot
Feature list: Informative Features
Optimization Metric: RMSE
Turn off time-aware modeling
What is the primary date/time feature?
Date DAYS
Mean Sales
2012-07-01
2014-06-15
Date
Time Series Modeling
What rolling window should DataRobot use to derive features?
28 to 0 days before Forecast Point
Which future values do you want to forecast?
1 to 7 days after Forecast Point
-28 days
0 days
+1 day
+7 days
PAST
Forecast Point (0 days)
FUTURE
Features known in advance: 0
Show Advanced Options
Explore Store Sales Forecast.xlsx


DataRobot
Data
Models 69
Deployments
Insights
Jupyter
Repository
NC_V6_New
Leaderboard
Learning Curves
Speed vs Accuracy
Model Comparison
Menu
Search
Add New Model
Filter Models
Export
Metric AUC
02
1
Model Name & Description
Feature List & Sample Size
Validation
Cross Validation
Holdout
Light Gradient Boosted Trees Classifier with Early Stopping
Tree-based Algorithm Preprocessing v1
M72 BP58 64.0%
RECOMMENDED FOR DEPLOYMENT
DR Reduced Features M55
80.0 %
0.6811 *
0.6786 *
0.6714
ENET Blender
M76 M59+61+55+70+56+...
MOST ACCURATE
Multiple Feature Lists
64.0 %
0.6844
0.6834
0.6751
AVG Blender
M74 M55+70+56
Multiple Feature Lists
64.0 %
0.6848
0.6828
0.6738
eXtreme Gradient Boosted Trees Classifier with Early Stopping
Tree-based Algorithm Preprocessing v20
M197 BP68
non-redundant features
64.0 %
0.6834
0.6817
0.6722
eXtreme Gradient Boosted Trees Classifier with Early Stopping
Tree-based Algorithm Preprocessing v20
M58 BP68
final features
64.0 %
0.6815
0.6815
0.6724

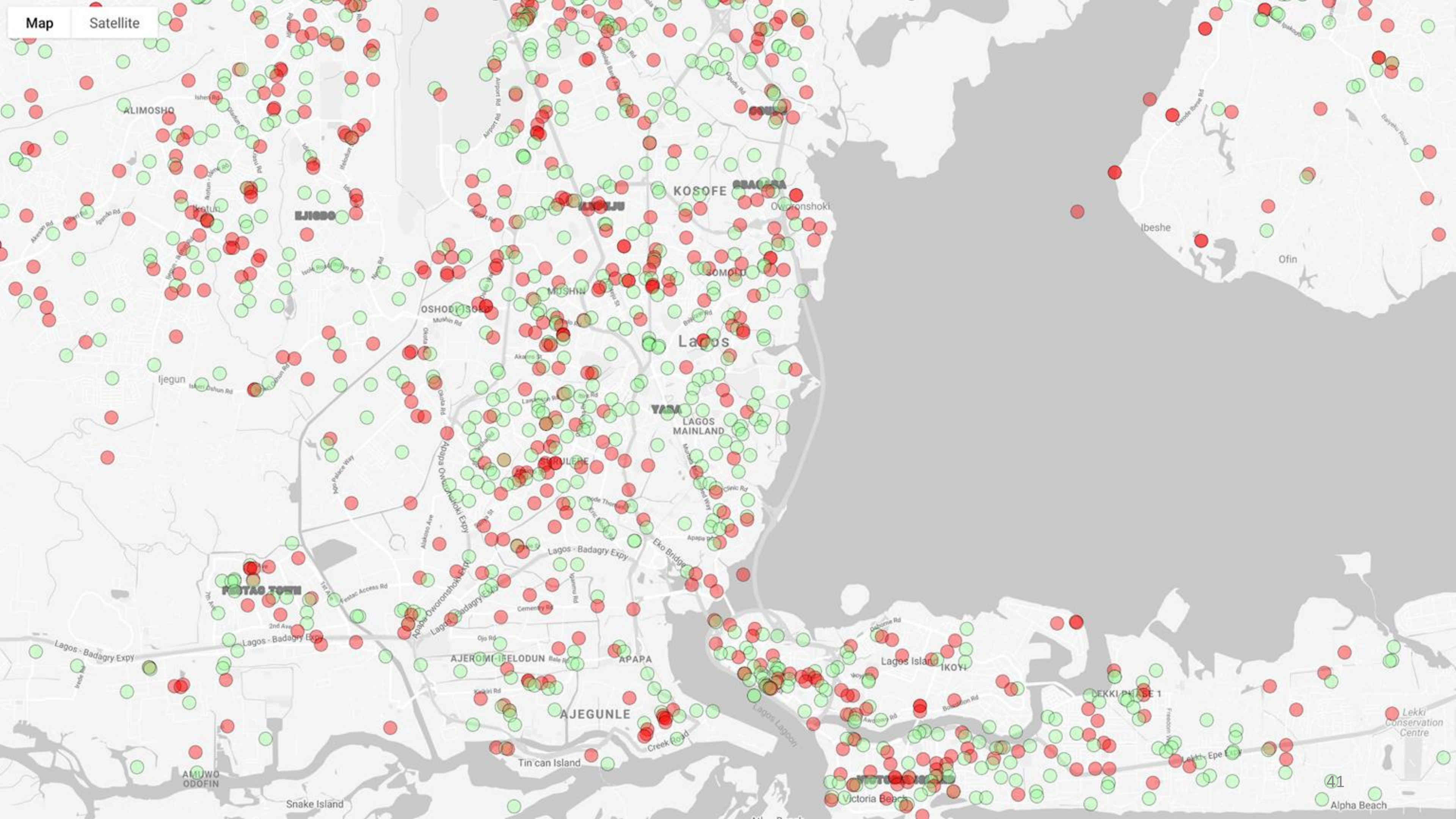

0.7x
Data
Model 60 Preprocessing
Model 70 Preprocessing
Model 58 Preprocessing
Model 56 Preprocessing
Model 59 Preprocessing
Model 61 Preprocessing
Model 57 Preprocessing
Model 55 Preprocessing
Gradient Boosted Greedy Trees Classifier with Early Stopping
Light Gradient Boosted Trees Classifier with Early Stopping
eXtreme Gradient Boosted Trees Classifier with Early Stopping
Gradient Boosted Trees Classifier with Early Stopping
eXtreme Gradient Boosted Trees Classifier with Early Stopping
Generalized Additive2 Model
eXtreme Gradient Boosted Trees Classifier with Early Stopping and Unsupervised Lear...
Light Gradient Boosted Trees Classifier with Early Stopping
Elastic-Net Classifier (L2 / Binomial Deviance)
Prediction

carbon

Pipelines Architecture



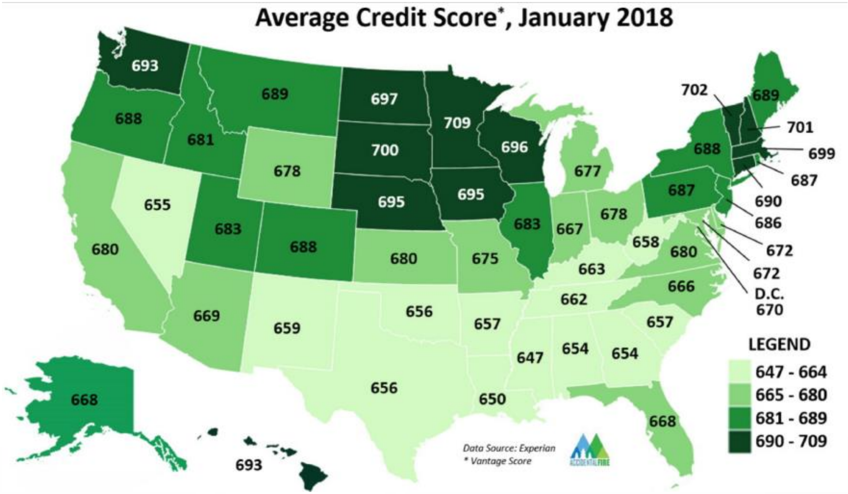
Why is automation essential ?

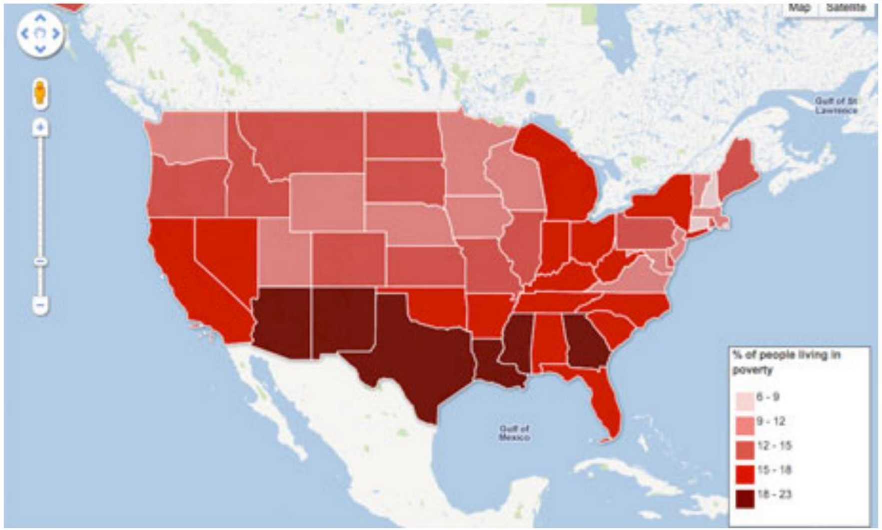

Map
Satellite
ALIMOSHO
Ikotun
EJIGBO
KOSOFE
Oworonshoki
Ibeshe
Ofin
SOMOLU
MUSHIN
OSHODI-ISOLO
Lagos
Ijegun
YABA
LAGOS MAINLAND
SURULERE
Eko Bridge
Lagos - Badagry Expy
FESTAC TOWN
AJEROMI-IFELODUN
APAPA
Lagos Island
IKOYI
LEKKI PHASE 1
AJEGUNLE
Tin can Island
Creek Road
AMUWO ODOFIN
Snake Island
Lagos Lagoon
Victoria Beach
Lekki - Epe Expy
Lekki Conservation Centre
Alpha Beach

7,344 GB
?

40 000 000 000 000 000 000 000 000

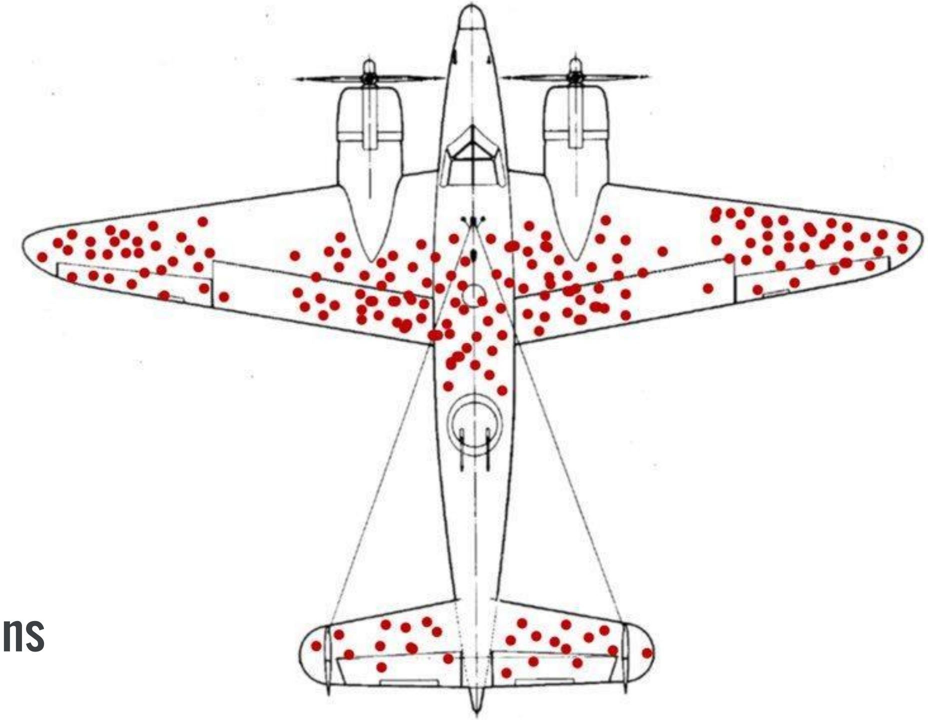
AI Bias vs Human Logical Error





US poverty interactive map. **Click image to explore it**

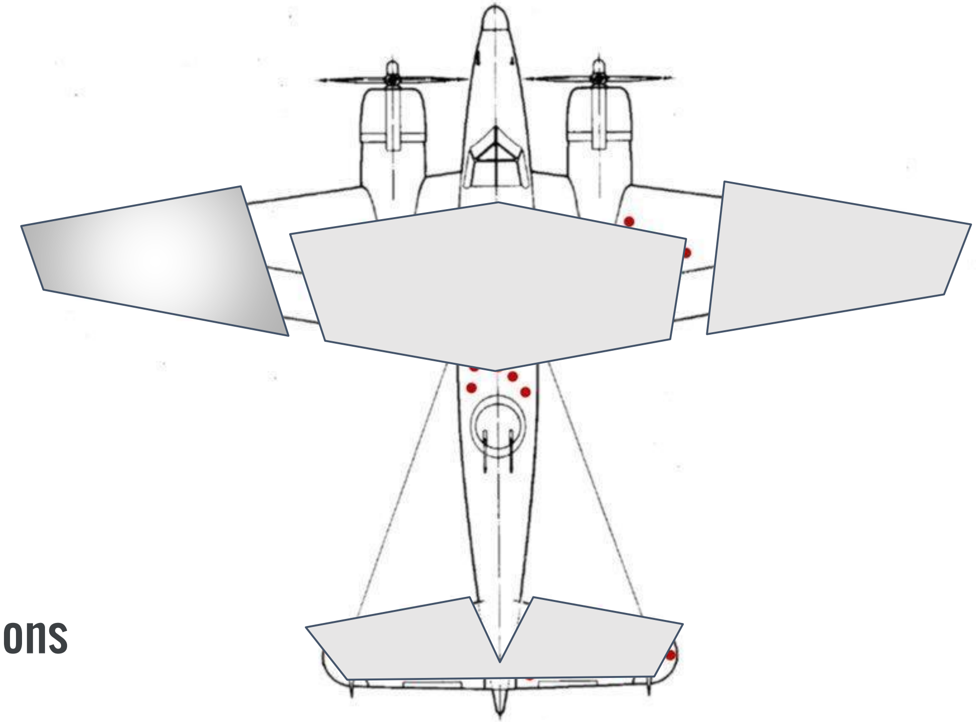
AI Bias vs Human Logical Error



Planes that returned from missions

AI Bias vs Human Logical Error

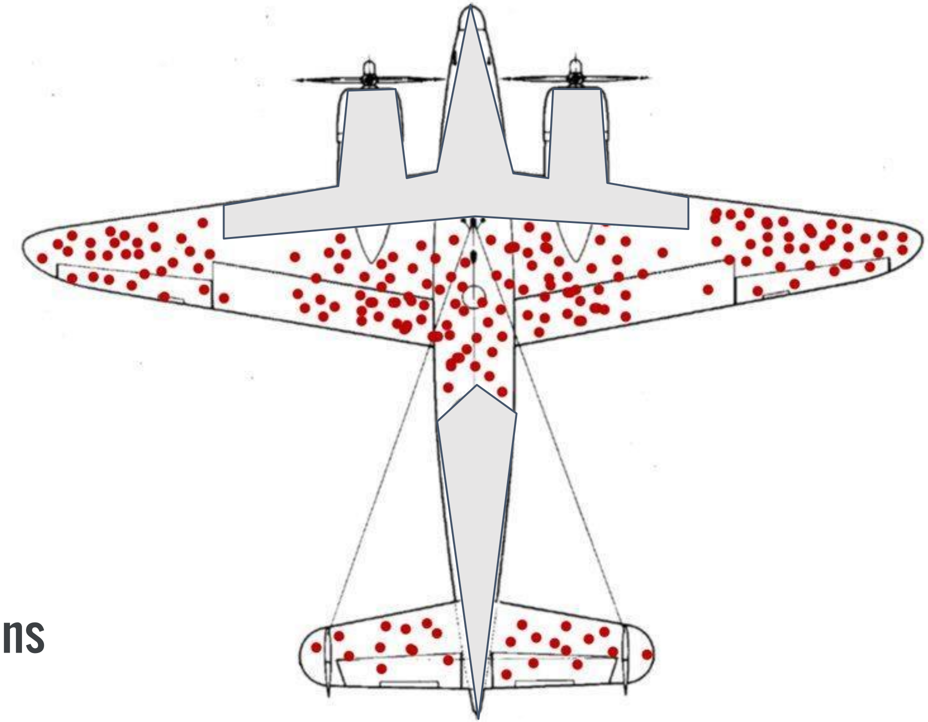
?



Planes that returned from missions

AI Bias vs Human Logical Error

=



Planes that returned from missions

Thank You